AD_____

Award Number: DAMD17-00-1-0353

TITLE: Microarray Technology to Study the Role of Genetic Polymorphisms in Breast Cancer Risk

PRINCIPAL INVESTIGATOR: Hilmi Ozcelik, Ph.D.

CONTRACTING ORGANIZATION: Mount Sinai Hospital
Toronto, Ontario, Canada M5G 1X5

REPORT DATE: July 2001

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2001	Annual (1 Jun 00 - 1 Jun 01)

4. TITLE AND SUBTITLE
Microarray Technology to Study the Role of Genetic Polymorphisms in Breast Cancer Risk

5. FUNDING NUMBERS
DAMD17-00-1-0353

6. AUTHOR(S)
Hilmi Ozcelik, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Mount Sinai Hospital
Toronto, Ontario, Canada M5G 1X5

E-Mail: Ozcelik@mshri.on.ca

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

It has been suggested that low penetrant alleles are associated with breast cancer risk. Although the contribution of low penetrant alleles to the individual risk is relatively small, they can contribute to a large proportion of breast cancer cases in the population. In this study we took the candidate gene approach to study the association of 32 different genetic polymorphisms in a population-based sample. We have also proposed to use the power of high-throughput microarray technology to identify low penetrant genetic polymorphisms which contributed to the risk of developing breast cancer. To date, we have successfully accomplished several goals in the development of this technology. Using the PCR primers designed for each polymorphism, a panel of cell lines has been screened using SSCP/Sequencing and the control DNA specimens corresponding to different alleles were determined. Allele-specific oligonucleotides for the purpose of probe preparation were designed for each allele, and using the control DNA samples and the PCR strategy their specificity were evaluated. A pilot SNParray was developed and several important factors were determined. Using this information, the full size SNParray containing all the proposed alleles were made. We are currently in progress to improve the analysis of the full size SNParray.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Cancer Genetics, Predisposition, genetic polymorphisms, microarray	12
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

It has long been hypothesised that genetic variation is responsible for observed differences in cancer risk and susceptibility amongst the human population. Mutant alleles of dominant highly penetrant breast cancer genes, including BRCA1 and BRCA2, do not occur frequently, and hence account for only a small proportion of breast cancer cases. On the other hand, several studies have suggested an association between low penetrant alleles and breast cancer risk. Although the contribution of low penetrant alleles to the individual breast cancer risk is relatively small, they can contribute to a large proportion of breast cancer cases in the population because the risk-conferring alleles of these genes are common.

Identification and cloning of low penetrant alleles that increase the risk of breast cancer is challenging because the association methods for such studies require large populations to achieve meaningful statistical analysis and very dense genetic maps to facilitate genome-wide genotyping. Although microarray technology has been developed to the point where it could be applied for parallel analysis of genome-wide genotyping, the dense genetic maps required for large population based association studies are currently being constructed for future genome-wide applications and will not be available for several years. At present, the candidate gene approach remains the most logical and practical strategy to identify these risk enhancing, low penetrant variants. Until now, a major obstacle with investigating the risk associated with multiple candidate genes has been a lack of technology for large-scale genotyping of large populations. Consequently, many studies have focused efforts on only 1 or 2 genetic polymorphisms, and even in these cases the analysis was only limited to relatively small sample sizes. Microarray technology is a solution to this obstacle. In the context of the ideas program, we propose to exploit the high throughput power of microarrays to simultaneously genotype 32 different genetic polymorphisms derived from 26 genes in a well defined, representative population-based sample containing a large number of subjects. We have selected genetic polymorphisms in genes involved in different aspects of carcinogenesis. For example, cell cycle regulatory genes such as CDK-inhibitors, and cyclins; carcinogen metabolizing enzymes such as CYPs, GSTs and NATs; immune system genes such as interluekins and TNF; and genes involved in other pathways involved in cancer (e.g. p53, PTEN, XPD-DNA repair gene). We have access to the Ontario Familial Breast Cancer Registry (OFBCR), which is the largest population based breast cancer registry in Canada. We also have support from the established microarray facility of the Ontario Cancer Institute in Toronto.

The objective of the proposed work is to identify low penetrant, yet commonly occurring, genetic polymorphisms which contribute to the risk of developing breast cancer. Furthermore, this approach has the potential to identify novel genetic factors associated with breast cancer risk, which may result in the development of innovative therapies, and a fuller understanding of genetic variation in response to therapy. The establishment of the proposed approach will prepare us for large-scale genotyping involving hundreds or even thousands of candidate genes in large define populations. This will lead to a more complex analysis of gene-gene and gene-environment interactions than is currently possible. Advances in disease etiology will significantly expand our abilities to design strategies for the prevention of breast cancer development and progression.

STATEMENT OF WORK

Task 1: Characterization of polymorphic alleles by SSCP, Months 1-8

a. Design of SSCP primers for 32 sites
b. Screen by SSCP analysis for all possible alleles at each locus
c. Sequence the SSCP patterns (appr 3 per loci) and identify the all possible genotypes

Task 2: Designing of oligonucleotides and sample microarrays, Months 4-8

a. Design different sets of oligonucleotides (perfect matches and mismatches)
b. Customize sample chips for quality control of hybridizations

Task 3: Optimization of the hybridizations using PCR probes, Months 8-16

a. Prepare PCR probes using control specimens
b. Optimize the hybridization conditions
c. Evaluate the accuracy of detection for every polymorphic site using a probes with different allelic combinations for each polymorphism
d. Redesign oligonucleotides and chips in order to increase the quality and accuracy of detection

Task 4: Genotyping of 900 specimens for 32 polymorphisms, Months 16-32

a. Production of microarray chips
b. Preperation of flourescent labelled PCR probes for each patient
c. Hybridization of chips at optimized conditions
d. Reading and analysis of the chip signals
e. Quality control experiments at different intervals using the control specimens to ensure the reproducibility of results

Task 5: Data and statistical analysis, Months 32-36

a. Repeat and conformation experiments
b. Complete the reading of every slide and prepare the data for statistical analysis
c. Univariate analysis of the data
d. Exploratory multivariate analysis of the data

BODY

1. Designing and Optimization of PCR Primers (Task 1)

Forward and reverse primers for all of 31 SNPs were designed using a computer program, Oligo4 (http://diogenes.baylor.edu). The primers were selected to amplify approximately 100-200bp sequences from the SNP regions of interest. The putative PCR region sequence for each SNP was compared to GenBank sequences using BLAST search (http://www.ncbi.nlm.nih.gov/). This is done to make sure that the gene sequences in interest did not share similar sequences in the genome (e.i. pseudogenes or gene family members). This comparison eliminated the possibility of such sequences contaminating our PCR product, which would have resulted in false detection of the allelic status.

Each pair of primers was optimized by titration of the $MgCl_2$ levels to determine the optimum concentrations of free Mg ions, thus providing a specific PCR product. These optimizations were done using DNA obtained from commercially available cell lines.

2. Identification of Heterozygous and Homozygous Control DNA Samples (Task 1)

After optimization of the PCR protocols of all the SNPs, a panel of cell lines were subjected to either SSCP or sequencing analysis. The aim of this part of the study was to determine heterozygous and homozygous DNA controls for each SNP which to be used as control samples during the microarray screening. All of 31 SNPs were screened and individual genotypes were identified. For SSCP experiments the PCR products were labelled using ^{33}P-dATP and the PCR products were analyzed on non-denaturing polyacrylamide gels containing glycerol. Sequencing reactions were performed using the ABI 377 automated sequencing system according to the manufacturers procedure.

3. Designing of the Allele-Specific and Support-Oligonucleotides (Task 2)

Allele specific oligonucleotides were designed for all of the 31 SNP regions using the Oligo4 computer program. For each SNP two allele-specific oligonucleotides were designed. Each allele-specific oligonucleotide had a unique TAG-sequence (~25 bases) attached to its 5'-end. These TAG-sequences are designed to be complementary to TAG-sequences (anti-TAG) on the support-oligonucleotides, which would be printed on the glass slides for hybridization purposes (described below). A gene specific region (~20 bases) is also attached to the TAG-sequence on both of the allele-specific oligonucleotides. The very 3'-nucleotide of each of these gene specific regions was designed to be complementary to the allelic nucleotides of each SNP.

Corresponding to each allele-specific oligonucleotide a support-oligonucleotide was also designed. Each support-oligonucleotides contained an anti-TAG sequence for each of the allele-specific oligonucleotide TAG-sequences to bind. Each of these anti-TAG-sequences were also attached to a 15-mer poly(T) tail. These support-oligonucleotides were printed on the glass slides using the poly-L-lysine surface chemistry (described below). The anti-TAG sequences on the support-oligonucleotides are designed to hybridize the flourescently labeled allele-specific oligonucleotides thus attaching them to the designated spots on the glass slides. The poly(T) tails on the other hand are designed to increase the efficiency of support-oligonucleotides to bind to the glass surface.

4. Measuring the Specificity of the Allele-Specific Oligonucleotides (Task 2 & 3)

The ability of each of the 62 (31x2) allele-specific oligonucleotides to specifically extend dNTPs was measured using allele-specific PCR strategy. Heterozygous and homozygous control DNA samples identified previously were used for this purpose. Each allele specific-oligonucleotide was amplified using a reverse primer specific to the same

SNP region. The quality and the quantity of each product were analyzed on agarose gels (Figure 1). The allele-specific oligonucleotides, which did not successfully amplified, were replaced with newly synthesized oligonucleotides. The oligonucleotides in good conditions were aliquoted and kept in freezer.

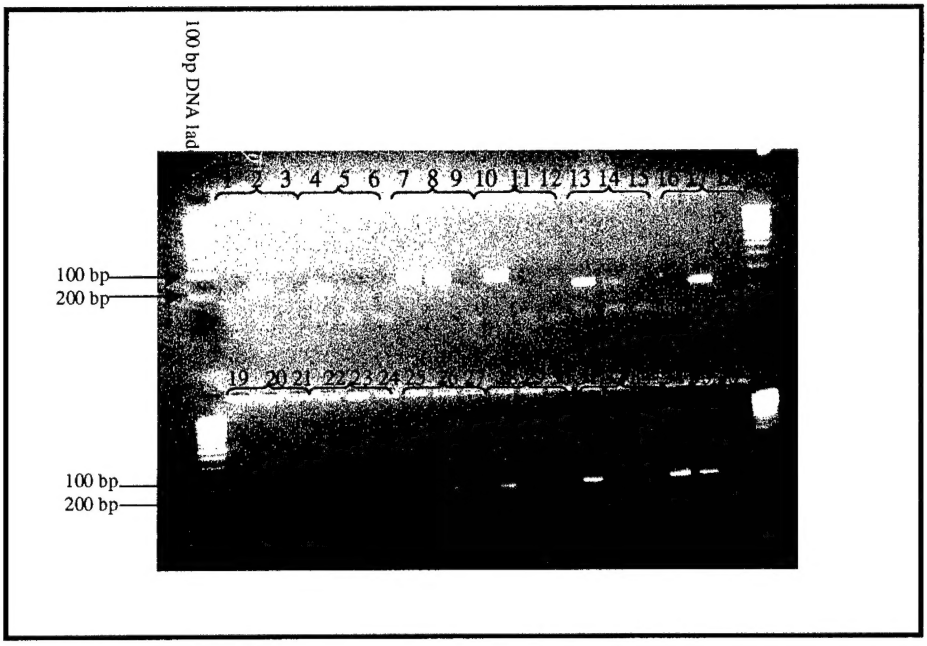

Figure 1 : Measuring the specificity and the quality of the extension of allele-specific oligonucleotides by allele specific PCR amplification. Each triplet shows the PCR reaction of two different DNA samples (heterozygous or homozygous) with one of the two allele-specific oligonucleotides of each SNP. They are Estrogen Receptor (ER) codon 10 C-allele, ER T-allele, ER codon 325, C-allele, ER G-allele, Matrix Metalloproteinase-1 (MMP1) insertion allele, MMP1 deletion allele, XPD T-allele, XPD G-allele, Glutathione S-Transferase 1 (STP1) T-allele, GSTP1 C-allele, GSTM3 A-allele, and GSTM3 G-allele specific oligonucleotides. The genotypes of the cell line DNAs used in these reactions were identified by sequencing previously. Third lane in each triplet shows the no DNA control reaction. Lanes 14 and 29 show unspecific binding and extension. First and last lanes of each row show the 100bp DNA ladder.

5. Pilot SNParray Study

In initial optimization studies, SNPs for Manganese Superoxide Dismutase (MnSOD) and Methlenetetrahydrofolate reductase (MTHFR) genes were utilized to determine several factors crucial for the sensitivity and the specificity of the SNParray strategy. These two SNPs were used in the determination of the concentration of the support-oligonucleotides to be printed on the glass slides, the surface chemistry of the glass slides to be utilized, the printing quality of the support oligonucleotides, and the hybridization conditions using the allele-specific PCR products. The design of the glass

slides using these two SNPs are shown in Figure 2A. A set of control support-oligonucleotides were also added to this design to specifically measure the printing quality of the oligonucleotides as well as the hybridization specificity (described below).

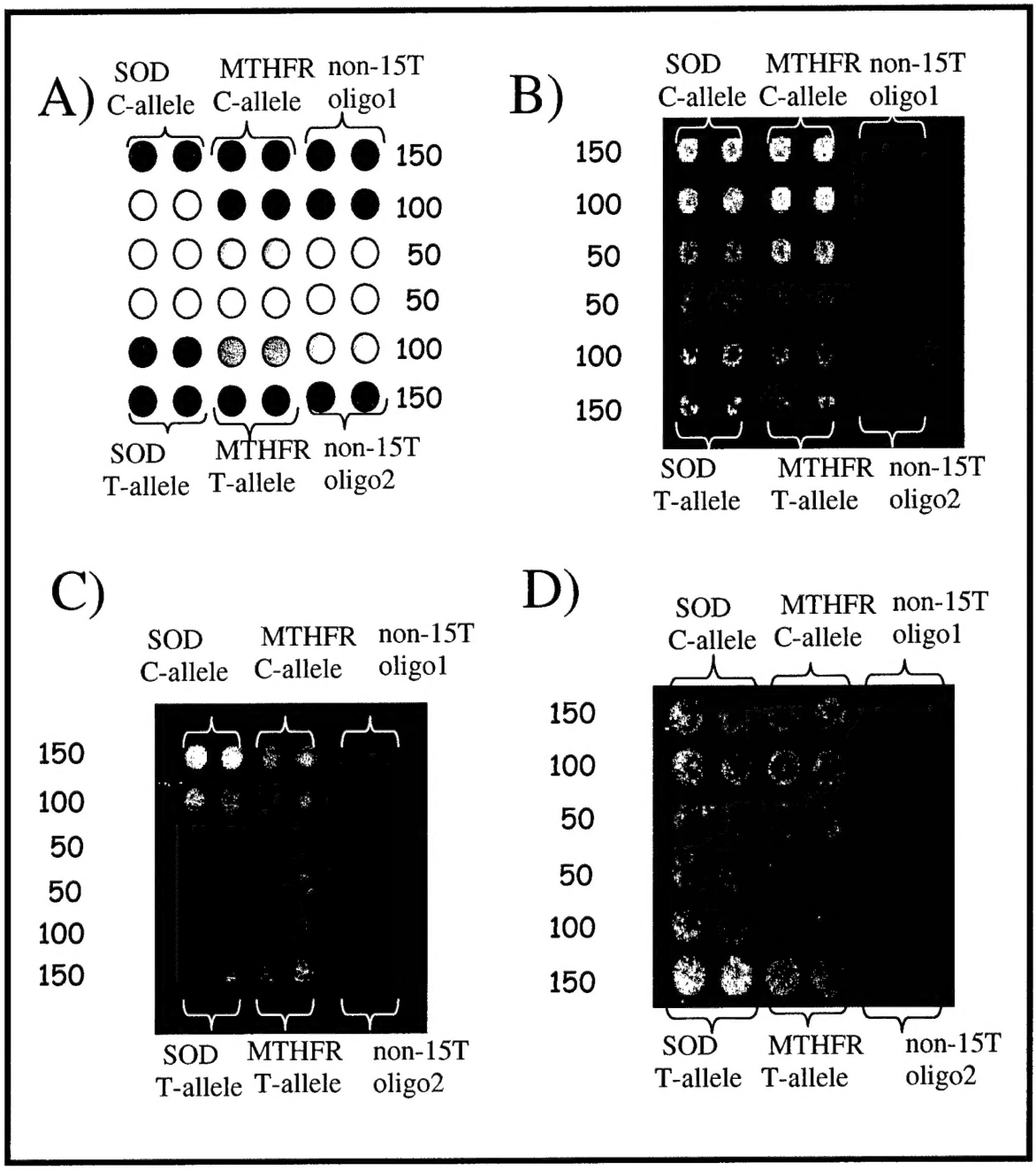

Figure 2: Design of pilot SNParray slide (A) and evaluation of the quality of oligonucleotide printing using super-aldehyde (B), super-amine (C) and Poly-L-lysine (D) surface chemistries. The hybridization reactions were done in the presence of CY3-labeled poly(A) control oligonucleotide probe, and the quality and the quantity of printing was evaluated. Each of the support-oligonucleotides was printed in three concentration, and each concentration was printed in duplicate.

5.A. Selection for the Surface Chemistry (Task 3)

The aim of this set of experiments was to determine the optimum surface chemistry of the microarray slides to be used for our strategy. Therefore, different slides and surface chemistries had been tried prior to selection.

Initially slides with super-aldehyde chemistry were used. For this purpose, support-oligonucleotides containing an amino residue at their 5'-ends were synthesized. The amino residues was designed to enhance the binding to the oligonucleotides to the glass slides from their 5'-ends leaving the rest of the oligonucleotide exposed to the probe. We have found that this type of chemistry provided high levels of fluorescent background, which did not help for the detection of the fluorescing spots in interest.

Super-amine surface chemistry unlike super-aldehyde chemistry provides a surface where negatively charged residues on the support-oligonucleotides can bind. These slides were covered with a positively charged chemical, an amine derivative. In order to enhance the strength of the attachment of the support-oligonucleotides on the super-amine coated surface a 15-mer poly(T) tail instead of a 6-carbon tail were attached to the oligonucleotides (as described above).

Slides with Poly-L-Lysine surface chemistry were also used. The mechanism of these slides to bind oligonucleotides is very similar to the super-amine slides. Again, the slides are covered with a positively charged chemical, Lysine in this case, for the negatively charged oligonucleotides to bind where the poly(T) tail enhances the strength of binding.

The printing quality of the three chemistries were evaluated using a CY3-labelled poly(A) oligonucleotide probe. All three types of slides containing MnSOD and MTHFR specific oligonucleotides as well as the control oligonucleotides lacking the poly(T) tail were hybridized to this probe and the results were evaluated using the laser scanner. The support oligonucleotides with complementary poly(T) tails (SOD and MTHFR) bound the flourescently labeled probe whereas the control oligonucleotide lacking this tail did not (Figure 2B-D). All of the three chemistries provided very satisfactory results in terms of the quality and the quantity of the printed oligonucleotides. However, we have selected to use the Poly-L-lysine chemistry since it provided strongest and the most specific hybridization signals during the course of hybridization with the allele-specific oligonucleotides for SOD and MTHFR (Figure 3).

5B. Probe Preparation and Hybridization

Allele-specific probes for each allele for SOD and MTHFR were prepared using an allele-specific and a reverse oligonucleotide. The PCR reaction was carried out in the presence of Cy5 labeled nucleotide using Taq DNA polymerase. Allele-specific oligonucleotides were extended using heterozygous and homozygous control DNA templates. The fluorescently labeled PCR probes were hybridized to the TAG sequences on the support-oligonucleotides printed on the slides. Hybridization was carried out in 2 X SSC + 0.1% SDS @ 50 ^{0}C for 3 hours. The slides were washed first in 2XSSC/SDS, 0.1% and then rinsed in 2XSSC solutions. The slides were dried in a centrifuge @ 500 rpm for 5 minutes. Hybridized slides were analysed for fluorescent signals in a laser scanner. Figure 3 shows the hybridization result for a homozygous and two heterozygous control DNA.

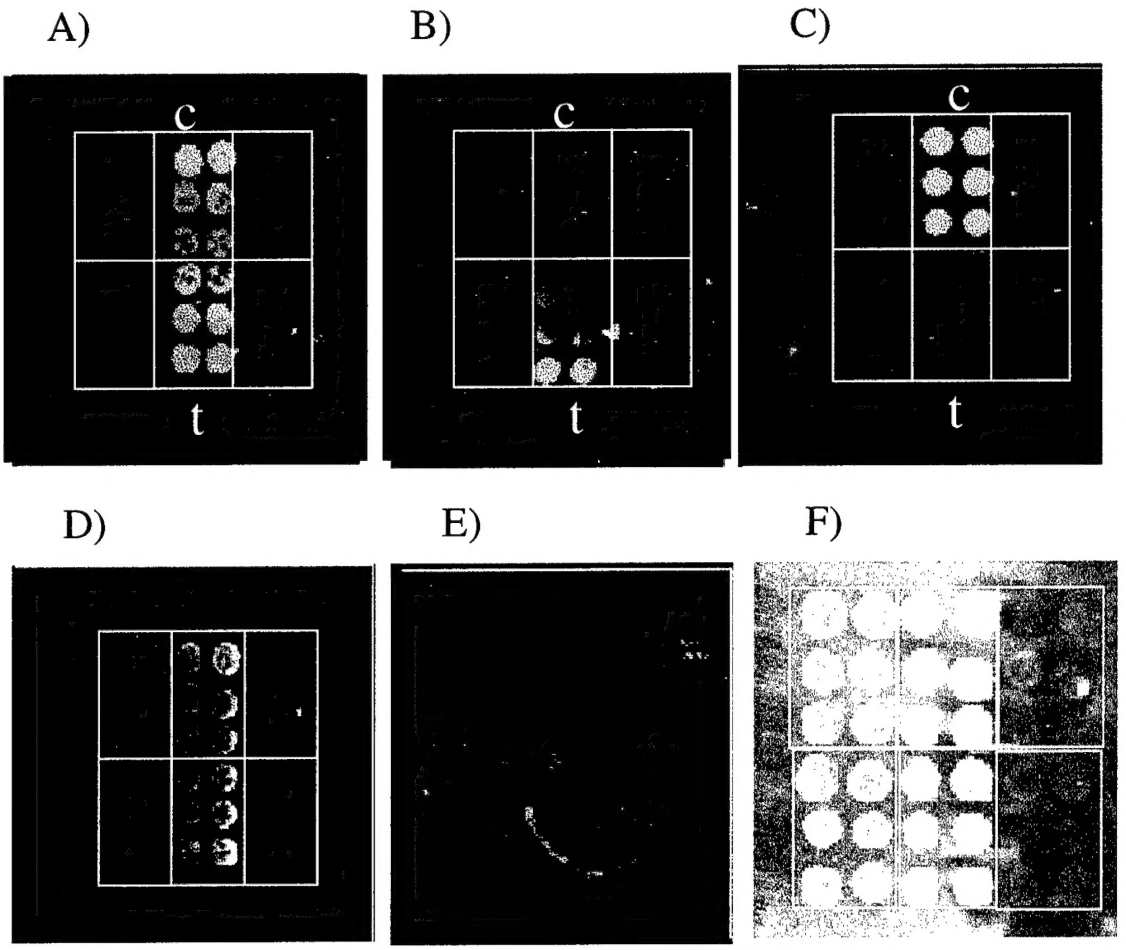

Figure 3: Genotyping of MTHFR SNPs using the optimized conditions described. Hybridization of reactions were done using heterozygous, ct (A), homozygous, tt (B) and homozygous, cc (C) control DNA samples. In each photo, no unspecific hybridization to either SOD or non15T control oligonucleotides were seen. Quality control of the printing for support oligonucleotides were done using the poly(A) tail probes prior to hybridization (D). Quality control hybridization gave signals on the SOD and MTHFR spots, but not on the non-15-T oligonucleotides. Cy-3 gives a visible green signal at 532 nm. Hybridization is also done using a DNA negative control performed during probe preparation for MTHFR hybridization reaction (E). Superimposition of Cy3-labeled poly(A) hybridization and Cy5 labeled primer extension reaction also done (F). When the slide is scanned in the ratio mode, both signals can be seen. When Cy3 and Cy5 are superimposed they give a yellow signal. Both SOD and MTHFR-TAG sequences give the green signal, indicating the poly(A) hybridization, but only MTHFR gives the red signal, indicating that the MTHFR allele-specific oligonucleotide extension. In this case, poly(A) hybridization also serves as the control of specific hybridization of MHTFR primer extension reaction to the MTHFR TAG sequences.

6. High Density Microarray Analysis with 62 Alleles

The proposed density of slides containing 62 different alleles corresponding to 31 SNP regions were printed on poly-L-Lysine slides. The design of the slide containing 61 SNPs is shown in Figure 4A. Each allele was printed in duplicate at 100uM concentration. Printing quality of these slides were also done several times to make sure that multiple slides within a batch provided reproducible results (Figure 4B).

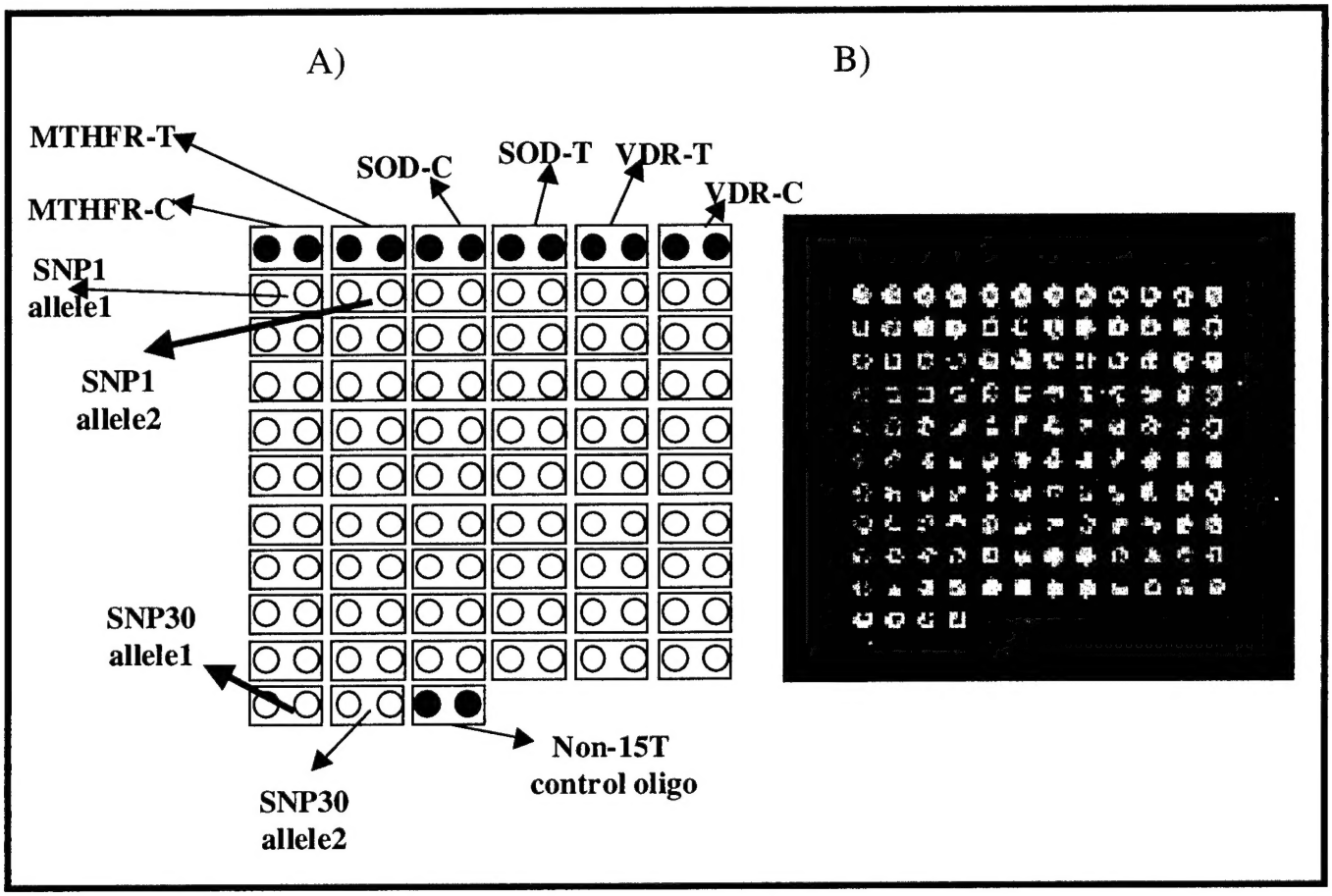

Figure 4: (A) The SNParray design with all 31 SNPs. B) Quality control for printing

7. Immediate Future Task

Using the full scale SNParray design we will evaluate the efficiency and the specificity of hybridization for each of the allele-specific PCR probes. Initially the PCR probes for each allele will be prepared separately and the efficiency and the specificity will be measured. After confirming that each and every allele-specific PCR probe is specific and, provided a good signal, then we will evaluate the specificity of using probes prepared by multiplex PCR strategy. We are currently in progress with these experiments.

8. Conclusion

We have accomplished the tasks proposed in the Statement of Work. We are now continuing to develop the method as proposed in the proposal.